LEVEL R

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
	AD-A100 994	

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
Optimal Outlier Tests for a Weibull Model - To Identify Process Changes or to Predict Failure Times	Technical Paper 7/1/80 - 6/1/81
	6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
Nancy R. Mann	N00014-80-C-0684, new N00014-76-C-0723

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Department of Biomathematics UCLA Los Angeles, CA 90024	NR047-204

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
Office of Naval Research, Code 434 Arlington, Virginia	May 1981
	13. NUMBER OF PAGES
	17

14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)	15. SECURITY CLASS. (of this report)
	Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

DTIC
ELECTE
JUL 7 1981
S D
A

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, If different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Weibull outlier tests, gap tests, identifying process changes

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper, Weibull outlier tests based on three different statistics are investigated with respect to their power optimality under various alternative models. Two of the statistics are new in the context of outlier statistics; and one of these is shown to provide a more powerful test in certain situations than other more classical outlier test statistics. Critical values of the three statistics were computer-generated and are tabulated. The tabulated values allow one to identify "treatment effects" resulting from unsuspected modifications to a process or to predict failure times in a life test. Numerical examples are given.

Optimal Outlier Tests For A Weibull Model -

To Identify Process Changes Or To Predict

Failure Times*

by

Nancy R. Mann

Department of Biomathematics

UCLA

Los Angeles, CA 90024

Abstract

In this paper, Weibull outlier tests based on three different statistics
are investigated with respect to their power optimality under various alter-
native models. Two of the statistics are new in the context of outlier
statistics; and one of these is shown to provide a more powerful test in cer-
tain situations than other more classical outlier test statistics. Critical
values of the three statistics were computer-generated and are tabulated.
The tabulated values allow one to identify "treatment effects" resulting
from unsuspected modifications to a process or to predict failure times in
a life test. Numerical examples are given.

1. Introduction

The results described in this paper pertain to the detection of Weibull outliers and to the prediction of a future ordered observation in an ongoing life test. The motivation for the research described herein, however, is the need for a method of determining whether or not, in a retrospective study, inordinately long times to failure are statistically significant and thus possible results of "treatment" effects caused by unsuspected modifications to a process.

Detection of outliers (spurious observations) is a problem that has long concerned experimenters and data analysts. An historical survey dealing with outliers was given as early as 1891 by Czuber [4]. A more up-to-date expository review of methods for detection of spurious observations was presented by Grubbs [10]. The latter paper is a modification of one "prepared primarily for the American Society for Testing Materials and represents a rather extensive revision of an earlier Tentative Recommended Practice ... ."

Grubbs points out that "almost all criteria for outliers are based on an assumed underlying normal (Gaussian) population" and Anscombe [1] in an extensive 1960 survey of the subject of outliers makes an initial assumption of normality for the data. (Discussion of the Anscombe paper and a paper by Cuthbert Daniel [5], dealing with outliers in factorial experiments, is given by William Kruskal, Thomas S. Ferguson, John Tukey, and E.G. Gumbel [15] and stresses the importance of the outlier problem.)

Most types of *life data* are such that a transformation cannot be made to impose normality on the underlying distribution. Thus, the traditional tests for and methods of treatment of outliers are inappropriate for most data arising from life tests. A statistic for testing for outliers in general location-scale families was recently proposed by Tiku [31] and shown to be more powerful

than various other statistics under Tiku's [32], p. 1418, outlier models ("labelled slippage" models of Barnett [2] and Barnett and Lewis [3]), although slightly less powerful under Dixon's [6] contamination models; see Tiku [31, 32], Hawkins [14] and Tiku [33], p. 139). The null distribution of Tiku's statistic is exactly Beta for the uniform and exponential populations and approximately Beta for the normal population (Tiku, [31, 32]); the percentage points are not available for any other distribution.

In the study described in the sequel, critical values were generated and have been tabulated for a variation of Tiku's statistic for a type-I extreme-value model (one in which the observations are logarithms of two-parameter Weibull variates). Critical values of two other statistics, shown under certain alternatives to be superior or essentially equivalent in terms of power, are also given. Analysis of optimality of power of tests is given in Section 3.3, and numerical examples are provided in Section 4.

2. Motivation

Often, during a life test, an experimenter has a need for an upper confidence bound (a prediction interval) for the time of the last (nth) failure in a size-n sample of test items. If the experimenter's data are two-parameter Weibull, Table 1 can be used to provide such a prediction interval for sample size $n = 5(1)25$, provided the first n-1, n-2, or n-3 failure times are known. On the basis of the first k failure times, with n-k=1, 2, 3, one can also use Table 1 to obtain an upper confidence bound for the time of the (k+1)st failure. By use of an approximation described in Section 3.2, it is also possible to obtain upper prediction bounds for the (j+1)st failure based on the first j failure times, with n-j=2, 3,..., n-2. This approximation can be applied for sample sizes ranging from 3 to as large as required.

Notwithstanding the usefulness of the results herein for obtaining certain prediction intervals, the primary motivation for the research described in the following was precipitated by analysis of data resulting from a large scale

retrospective longitudinal study of times of individuals relapsing to undesirable habitual behavior. Results of Mann and Rothberg [26] and Mann [21, 22] appear to indicate that either a two-parameter Weibull model or a mixture of two-parameter Weibulls is appropriate for "time-to-failure" or return to addictive or other undesirable habitual behavior for longitudinal studies made on individuals. Here, it is convenient to conceptualize independent intentions to abstain from the behavior that wear out or otherwise fail in time. (Time-to-first failure in a cohort has been studied in the case of prison recidivism by Harris and Kaylan [13], who found that a mixture of two exponentials provided a good fit for the data.)

What one is attempting to determine in applying an outlier test to retrospective longitudinal time-to-failure data is whether or not "treatment effects" may have resulted in specified instances. If the Weibull outlier test indicates that a number of seemingly inordinately long times to failure are significantly different from other failure times of an individual, then one can attempt to correlate the instances involving suspected treatment effects with various potential causal factors.

Such an outlier test can be used, as well, to identify treatment effects in hardware on the basis of life-test data. In such situations, identification of an outlier will potentially allow one to discover inadvertent and/or unsuspected modifications that may have been made to a manufacturing process. Note that the immediate goal is not parameter estimation, as in many situations, and also that rather large numbers of outliers are a definite possibility.

3. Determination of Appropriate Test

3.1 Earlier Results

Tiku [31] defined $\hat{\sigma}$ to be the (size-n) maximum-likelihood estimator of the scale parameter of a location-scale-parameter distribution (i.e., a

distribution $F_X(x)$ that is of the form $G[(x-\mu)/\sigma]$ for some G). He defined σ_c to be the maximum-likelihood estimator of σ, or an estimator with the asymptotic properties of the maximum likelihood estimator of σ, calculated from all the k<n ordered observations felt not to be outliers (considered together as a censored size-n sample); i.e., σ_c is consistent, asymptotically unbiased and efficient and asymptotically normal for the cases he considered and for the case considered here.

Tiku then proposed

$$T = h(\sigma_c/\hat{\sigma}) \qquad (3.1.1)$$

(where h is a suitable constant) as a statistic for testing the hypothesis that the sample contains no outliers versus the hypothesis that the suspect observations are all outliers. He demonstrated empirically, for 1, 2 and 4 outliers, n=10, 20 and 40, that the statistic T has higher power than certain other well known statistics (see Grubbs [10], Tietjen and Moore [37], Shapiro and Wilk [30] and Ferguson [9]) under Tiku's [31, 32] labelled slippage models (Models A and B of Section 3.3). Note that Tiku's statistic is versatile; (i) it can be used to test any specified number of outliers on either side of an ordered sample, and (ii) it can be used to test whether the sample contains outliers, irrespective of how many [32]. p. 1420. A multivariate generalization of Tiku's statistic is also available (Tiku and Singh, [35]).

Outliers on the left are not generally of interest in our analyses. They often arise because inspections of hardware or tests for abstinence (such as urinalyses to test for opiates and other drugs) are made at discrete time intervals, perhaps weekly. Thus, small values are relatively more displaced than larger values. Because of the logarithmic transformation, any displacement of small values is magnified as well.

Now, consider a sample with a single large suspected outlier from a one-parameter exponential distribution with parameter σ. Here σ_c and $\hat{\sigma}$

are equal to $S_{n-1}/(n-1)$ and S_n/n, respectively, where

$$S_j = \sum_{i=1}^{j} X_{(i)} + (n-j)X_j,$$

with $X_{(i)}$ the ith exponential order statistic. Thus, for this distribution (in which σ is both a location and scale parameter), the statistic T is proportional to $(n-1)\sigma_c/(n\hat{\sigma}) = S_{n-1}/S_n$, which is equal to $S_{n-1}/[S_{n-1} + (X_{(n)} - X_{(n-1)})]$. If U_k is defined to be $(X_{(n)}-X_{(k)})/S_k$, then $(n-1)\sigma_c/(n\hat{\sigma}) = (1 + U_{n-1})^{-1}$ in this single outlier case.

Lawless [16] proposed the use of U_k for obtaining a prediction interval on $X_{(n)}$, the nth ordered observation, from the first k observations in a life test in which the data are exponential with parameter σ; and he demonstrated that for (one-parameter) exponential data, $(n-1)$ U_{n-1} is distributed as Snedecor's F with 2 and 2n-2 degrees of freedom.

Monte Carlo results exhibited in Table 3 demonstrate similarly that for data from an extreme-value distribution (data that are ordered logarithms $(X_{(1)}<\cdots<X_{(n)})$ of sample observations from a two-parameter Weibull distribution), the power of a test based on T is equivalent to the power based on the ratio of $(X_{(n)} - X_{(n-1)})$ and an estimate equivalent to the maximum likelihood estimate of the extreme-value scale parameter (the Weibull shape parameter) obtained from the first n-1 observations.

For more than a single large outlier, the statistic T defined above involves observations that are not available in the prediction interval situation. Hence, for any distribution, using a statistic similar to U_k, i.e., proportional to $Q_{\ell-k} = (X_{(\ell)} - X_{(k)})/\sigma_c$, $k < \ell \leq n$, for testing for n-k outliers would seem to be inefficient for n>k+1. It will be shown in Section 3.3 that this is not necessarily so.

3.2 Test Statistics for Weibull Data

We consider now the variate X, the logarithm of a Weibull variate with

$$F_X(x) = \begin{cases} 1 - \exp[-\exp\{(x-\mu)/\sigma\}], & x > 0 \\ 0, & \text{otherwise}; \quad \sigma > 0 . \end{cases}$$

The parameter μ is a location parameter, the mode of the distribution of X (the first asymptotic distribution of the smallest extreme) and is the logarithm of the Weibull scale parameter. The parameter σ, which determines the shape of the Weibull distribution, is a scale parameter of the distribution of X, with $\pi^2\sigma^2/6$ the variance of X.

Since X has a location-scale parameter distribution, it is to be expected that for the labelled slippage model of Tiku (see Section 3.3), an efficient test statistic for testing for large outliers can be provided by $T = h(\sigma_c/\hat{\sigma})$. One might also consider statistics proportional to $Q_{\ell-k}$, $k < \ell \leq n$.

Results of Lawless [16], Thoman, Bain and Antle [36], and Mann and Fertig [23], show that for Weibull data, maximum-likelihood and best linear invariant estimators yield very nearly equal numerical results and their small- and large-sample properties (bias, mean squared error, etc.) are very nearly equivalent. Thus, for testing that the largest n-k of n sample observations are outliers, using T is essentially equivalent to using as a test statistic $\tilde{\sigma}_{k,n}/\tilde{\sigma}_{n,n}$, the ratio of the best linear invariant estimators of σ based on the smallest k and on all n sample observations, respectively. The power is obviously unchanged if one uses $\sigma^*_{k,n}/\sigma^*_{n,n}$, the ratio of the best linear unbiased estimators of σ based on the smallest k and on all n sample observations respectively. This is true since best linear invariant and best linear unbiased estimators of σ differ only by a constant factor. See, for example, Mann [19].

In this study we considered specifically

$$V_{n-k} \equiv hT^{-1} = \tilde{\sigma}_{n,n}/\tilde{\sigma}_{k,n} \, ,$$

$$Q_{n-k} = (X_{(n)} - X_{(k)})/\tilde{\sigma}_{k,n}$$

and

$$W_{n-k} = Q_{(k+1)-k} = (X_{(k+1)} - X_{(k)})/\tilde{\sigma}_{k,n} \qquad .$$

Note that Q_{n-k} and W_{n-k} yield gap tests somewhat similar to some suggested by Dixon [6]. Critical values of these statistics for testing for large outliers, or predicting later failure times, at 0.20, 0.10, 0.05 and 0.01 significance levels for n = 5(1)25, n-k = 1, 2, 3, are displayed in Table 1, and an example of their use is given in Section 4.

The values shown for V_{n-k} and Q_{n-k} were generated simultaneously by means of 20,000 Monte Carlo simulations. The exhibited values of W_{n-k} were generated by making use of the fact that, for $k \leq n - 2$ (the restriction having been discovered in this research),

$$F_k = [(X_{(k+1)} - X_{(k)})/E(X_{(k+1)} - X_{(k)})]/\tilde{\sigma}_{k,n}/E(\tilde{\sigma}_{k,n})$$

has approximately a classical F distribution. This is discussed in Mann, Schafer, Singpurwalla [27], pp. 255-256.

In order to generate the tabulated values of W_{n-k}, using the F approximation, it was necessary to use stored values of the expectations of the reduced order statistic $Y_{i,n} = (X_{(i)} - \mu)/\sigma$, i=k, k+1, and of $C_{k,n}$, where $\sigma/(1 + C_{k,n})$ is expectation of $\tilde{\sigma}_{k,n}$ and $C_{k,n}\sigma^2$ is the variance of

$$\sigma^*_{k,n} = (1 + C_{k,n})\tilde{\sigma}_{k,n} \, ,$$

the best linear unbiased estimator of σ, based on the smallest k observations of X. Thus,

$$F_k = (X_{(k+1)} - X_{(k)})/[E(Y_{k+1,n}) - E(Y_{k,n})]/[\tilde{\sigma}_{k,n}(1 + C_{k,n})] \, .$$

The degrees of freedom for the approximate variate are based on the result of Patnaik [24], which specifies for ϕ, with $E(\phi) = m$, $\mathrm{var}(\phi) = v$,

and m^2 proportional to v, that $2m\phi/v$ is approximately a chi-squared variate with $2m^2/v$ degrees of freedom. Thus, we have for F_k,

$\nu_1 = 2 \ var[(Y_{k+1,n} - Y_{k,n})/E(Y_{k+1,n} - Y_{k,n})] \sim 2$ and $\nu_2 = 2/C_{k,n}$ degrees of freedom. Values of $var(Y_{k+1,n} - Y_{k,n})$, n-k = 2,3; n = 5(1)25, were calculated from stored values, along with the other constants needed for the compu- tations. (See below for the origin of these constants.)

The values obtained from the F approximation were compared with trial simulations having a Monte Carlo sample size of 20,000 to ensure that the tabulated values are sufficiently precise. The agreement increases as significance level α decreases. That is, higher percentile values are more precise. Also, precision increases as sample size n increases and as k decreases. Examples of comparison with Monte Carlo values are shown in Table 4.

The method used for obtaining the F values with noninteger degrees of freedom is described in Mann, Schafer, Singpurwalla [27], pp. 172, 173. This method, along with values of

$E(Y_{k+1,n} - Y_{k,n})$,   k = 2, n-1,

tabulated in Mann, et al. [27] pp. 342-347 for n = 3(1)16, and Mann, Schauer and Fertig [28], for n = 3(1)25, and values of $C_{k,n}$, which can be obtained from values appearing in Mann, et al. [27], pp. 194-207, for n = 2(1)13 and in Mann [19], for n = 2(1)25, can be used to estimate the critical values of W_{n-k} for n-k>3. In these cases, one can use $\nu_1 = 2$ along with $\nu_2 = 2/C_{k,n}$ for the degrees of freedom or can calculate ν_1 more precisely using values of the variances and the covariance of $Y_{k+1,n}$ and $Y_{k,n}$ available in Mann [20].

For samples larger than 25 and n-k>1, one can use the approximation with asymptotic expressions for expectations, variances and covariances of the order statistics available in Mann et al. [27], p. 218, and an asymptotic expression for $C_{k,n}$ available in Harter and Moore [12].

As noted earlier, maximum-likelihood estimates can be substituted for $\tilde{\sigma}_{k,n}$ and for $\tilde{\sigma}_{n,n}$, and the values in Table 1 can be used directly with these estimates without any modification required. One can also use $\sigma^*_{k,n}$ and $\sigma^*_{n,n}$, best linear unbiased estimates (see Mann [19]) or simplified linear estimates (see Mann, et al. [27], pp. 210-212, Mann and Fertig [24] or Engelhardt and Bain [7,8]), in place of the best linear invariant estimates. In this case, the modified statistics Q_{n-k} and W_{n-k} need to be multiplied by the factor $CQW = (1 + C_{k,n})$ and the modification of V_{n-k} needs to be multiplied by $CV = (1 + C_{k,n})/(1 + C_{n,n})$ before comparison with critical values. In other words, $\sigma^*_{k,n}$ and $\sigma^*_{n,n}$ need to be divided by $(1 + C_{k\,n})$ and $(1 + C_{n,n})$ respectively, to convert them to $\tilde{\sigma}_{k,n}$ and $\tilde{\sigma}_{n,n}$. Values of the constants CQW and CV appear in Table 2 for $n = 5(1)25$, $n-k = 1, 2, 3$.

Approximations to Q_{n-k} and V_{n-k} can be calculated by using probability plots such as those shown in Figures 1 and 2. Here, the inverse of the slope of the line plotted on the basis of the smallest k observations gives an approximation to $\sigma^*_{k,n}$; and the inverses of the slopes of the line formed by the nth and k^{th} points and by the line formed by the (k+1)st and k^{th} points give approximations to $(X_{(n)} - X_k)/E(Y_{n,n} - Y_{k,n})$, and $(X_{(k+1)} - X_k)/E(Y_{k+1,n} - Y_{k,n})$, respectively.

If the inverse of the slope in the probability plots is used, then the constant factor

$$CQP = E(Y_{n,n} - Y_{k,n})(1 + C_{k,n}) \quad \text{or}$$

$$CWP = E(Y_{k+1,n} - Y_{k,n})(1 + C_{k,n}) \ ,$$

must be used to multiply the value obtained to convert it to one that can be compared with the critical factors for Q_{n-k} or W_{n-k}, respectively. Values of CQP and CWP are given in Table 2 for $n = 5(1)25$; $n-k = 1, 2, 3$.

Special probability papers, each one applicable to a specified sample size, have been designed (see [18]), so that individuals without technical training can plot failure times of interest. Without making such plots, one will usually find it very difficult to have much feeling for what might be moderately large values for time-to-failure when the data are Weibull. Using the plots with some minimal instruction, a nontechnical person should be able to determine slopes of lines formed by $x_1,\ldots,x_k$ and by x_k and x_n or x_k and x_{k+1}. This assists a spouse, a "significant other" or a counsellor of a subject engaging in undesirable habitoral behavior to gain insight into what might be, for this subject, motivation for long-term abstinence.

3.3 Optimality of Power Under the Two Alternatives

For a Weibull model, the hypothesis H_o to be tested is:
$X_{(1)},\ldots,X_{(n)}$ are order statistics from

$$f_X(x) = \frac{1}{\sigma} \, g[(x-\mu)/\sigma] \qquad\qquad (3.3.1)$$

where $f_X(x)$ is the density function corresponding to the distribution function (3.2.1). Model A and Model B are given, respectively by

A: $X_{(1)},\ldots,X_{(k)}$ are the smallest k order statistics from (3.3.1) and

$X_{(k+1)},\ldots,X_{(n)}$ are the largest n-k order statistics from

$$f_X(x) = \frac{1}{\sigma} \, f\{[x - (\mu + \delta\sigma)]/\sigma\}$$

and

B: $X_{(1)},\ldots,X_{(k)}$ are the smallest k order statistics from (3.3.1) and

$X_{(k+1)},\ldots,X_{(n)}$ are the largest n-k order statistics from

$$f_X(x) = \frac{1}{\lambda\sigma} \, g[(x - \mu)/\lambda\sigma] \; .$$

These models may not correspond to the manner in which data are generated for the situation described. Nonetheless, a mixture of any two specified

Weibull distributions can be represented by a mixture of models A and B if the "outliers" are larger than other values and the number of outliers is only one or two. Models A and B can be combined also to approximate very well nearly any model that is a mixture of a Weibull sample of small values and a Weibull sample of larger values (the "outliers").

Examples of Model A and Model B are shown as probability plots (on Weibull probability paper) in Figures 1 and 2, respectively. It was the object of the research described in this paper to determine test statistics that are optimal, in terms of power considerations, for testing for outliers, in general, and for testing against Model A or Model B, or a mixture of these, in particular. To this end, the power of the various test statistics under consideration was calculated by 2000 Monte Carlo simulations (in addition to the 20,000 used to generate critical values for the test statistics). These power calculations were made for each critical value generated for V_{n-k}, Q_{n-k} and for selected sample sizes for W_{n-k} for Model A: $\delta = 0.5, 1$. Model B: $\lambda = 2, 5$ and mixed models $\delta = 1$; $\lambda = 2, 5$. Illustrative examples are exhibited in Table 3. Note that only for the test statistic W_{n-k} (under Model A with $n \geq 10$) does the power increase as n-k increases. This is probably due to the fact that observations near to μ are closer together than observations near the tail. Hence, displacement of 1 σ is less critical near the tail.

On the basis of the many similations that were made, it has been well established that when one is testing H_0 versus a single outlier, a test based on $Q_{n-k} \equiv Q_{n-(n-1)} \equiv W_{(k+1)-k} \equiv W_{n-(n-1)}$ has power essentially identical to that of one based on $V_{n-(n-1)}$ hT^{-1}. This was pointed out in Section 3.1.

Tiku [27] has demonstrated for Gaussian families that a test based
on the statistic T has higher power in more general situations (more than
a single outlier) than other classical outlier tests under his labelled
slippage model. As the number of outliers, n-k, increases, however, the
ratio of the power of W_{n-k} relative to the power of V_{n-k} increases under
Model A (shift in location). That is to say, under Model A, a test based
on a measure of the gap $(X_{(k+1)} - X_{(k)})$ between the smallest suspected
outlier and the largest observation thought not to be an outlier, relative
to a measure of the dispersion $(\tilde{\sigma}_{k,n})$ of the observations thought not to
be outliers is more powerful than one based on T (see Table 3). It is
clear from Figure 1 that for Model A, it is essentially this quantity,
i.e., the size of the gap relative to the dispersion of the smaller obser-
vations, that is the critical factor in establishing the suspicion of
outliers. Thus, it is not unlikely that a test based on a statistic, such
as W_{n-k} involving $X_{(k+1)} - X_{(k)}$, is optimal for alternative models resem-
bling Model A.

If it were established that Model A was precisely the alternative
(which it usually will not be), then using in the denominator of W_{n-k}
an estimator of σ that involves all differences of successive order statistics
except $X_{k+1} - X_k$ would be more powerful than W_{n-k} as it is defined. Such
a test would be equivalent in terms of power, to one having this statistic
in the denominator and $\bar{\sigma}_{n,n}$ in the numerator and should be optimal for the
labelled slippage model with Model A as the alternative. Note that Mann
and Fertig [25] demonstrate that for a goodness-of-fit test, involving
gaps (which all estimates of σ in location-scale families involve) the
important consideration in determining optimality in which gaps are in-
volved in the test and in what position, rather than how the gaps are
combined. That is to say, an optimal estimator of σ based on the first

k-1 gaps performs no better than one which is the sum of each of the k-1 gaps divided by its expectation.

In this context it is noted that the statistic $V^* = \tilde{\sigma}_{k+1,n}/\tilde{\sigma}_{k,n}$ for $k = n-2, n-3, \cdots$, has the same functional relationship with W_{n-k} that $V_{n-(n-1)}$ has with $Q_{n-(n-1)}$. Therefore, the statistic W_{n-k} also has essentially the same power as V^*. The inverse of V^* is a special case of Z, a statistic proposed by Tiku [34, eq. 1.4] for testing goodness of fit when H_0 is exponentiality. The statistic Z is equivalent to V^* when the exponential censored sample of size n consists only of the smallest k+1 observations. Thus, Z stresses the difference of the two largest observed order statistics.

It should also be pointed out (see [24]) that μ^* and $\tilde{\mu}$ based on $X_{(1)}, \cdots, X_{(k)}$, k<n, from an extreme-value distribution are of the approximate form, $X_{(k)} + c\tilde{\sigma}_{k,n}$, where c is an appropriate constant. Thus, a test of form $(X_{(k+1)} - \tilde{\mu}_{k,n})/\tilde{\sigma}_{k,n}$ is essentially the test W_{n-k}.

For Model B, the critical factor is the ratio of the slopes of the plots of the smallest k and the largest n-k+1 observations. For this model, W_{n-k} performs poorly relative to V_{n-k}, as one might suspect, but Q_{n-k}, which is proportional to the ratio of estimates of these two slopes, approximates V_{n-k} very well, i.e., powers of V_{n-k} and Q_{n-k} are very nearly equivalent. See Table 3. Thus statistic Q_{n-k} is shown (in Table 3) to perform very poorly, in terms of power, under Model A, however.

For a mixture of the models, results shown in Table 3 indicate that a test based on W_{n-k} tends to be most powerful, with Q_{n-k} performing most poorly. Again, as with Model A, the gap $X_{(k+1)} - X_{(k)}$ relative to $\tilde{\sigma}_{k,n}$ appears to be the most critical factor.

It seems clear from this study that in considering whether or not to test for outliers, one should, if possible, plot the data on probability paper. Plotting is useful in providing perspective even though there is a single suspected outlier. For more than a single outlier, plotting is essential if one is to know whether to use W_{n-k} (for Model A or mixed models) or either V_{n-k} or Q_{n-k} (for Model B) or V_{n-k} (for a more general alternative model). In this way no one can insure using a test with what appears to be optimal power.

Clearly, the power of the outlier test is affected by the apriori analysis, as is always the case to some extent in looking at the data before performing an outlier test. However, in this context it is important to identify large outliers in order to determine if treatment effects (extending life or for human subjects, extending periods of abstinence) have resulted and what might have caused such effects. The goal is not primarily one of estimation of parameters, but rather of exploration. This point is discussed by Barnett and Lewis [3], pp. 5-6.

Finally, it is to be noted that the results obtained here are likely to extend to other location-scale families. Thus, an analog of W_{n-k} involving the gap $X_{(k+1)} - X_{(k)}$ will possibly tend to be more powerful for any location-scale family (including Gaussian distributions) for testing H_o under Model A than is the statistic T.

4. Numerical Examples

The data in the probability plots (Figures 1 and 2) are used here to provide examples of the use of the various test statistics.

First, we consider Figure 1, which exhibits two possible outliers from a mixed model with $\lambda < 1$. Here n is equal to 9, so that tables in [22] can be used to obtain the weights to calculate $\tilde{\sigma}_{7,9} = 0.709$ and $\tilde{\sigma}_{9-9} = 0.884$. Also $X_{(9)} - X_{(7)} = 0.872$ and $X_{(8)} - X_{(7)} = 0.693$. Thus, $v_{9-7} = 1.245$, $q_{9-7} = 1.228$

and $w_{9-7} = 0.976$. Comparing these values with the tabulated critical values, one finds that if the specified significance level is 0.10, then only the test statistic w_{9-7}, involving $x_{(8)} - x_{(7)}$, rejects the hypothesis of no outliers.

The plotted line drawn (by hand) in Figure 1 gives highest weight to the kth, or in this case, the seventh value, as do the weights for optimal linear estimates of σ, such as $\tilde{\sigma}$ and σ^*. Also note that horizontal, rather than vertical, distances from points should be minimized. The slope of the line is about 1.20 so that an approximation to $\sigma^*_{7,9}$ is about 0.833. This gives 0.717 as an approximation to $\tilde{\sigma}_{7-9}$ with the use of CQW = 1.161 (found in Table 2) as a divisor.

The plot in Figure 2 suggests 3 large outliers of the general type specified by Model B. Thus, using tabulated values in [14], one finds $\tilde{\sigma}_{11,14} = 0.84226$, $\tilde{\sigma}_{14,14} = 1.353$, $x_{(14)} - x_{(11)} = 2.084$ and $x_{(12)} - x_{(11)} = 0.560$ so that $v_{14-11} = 1.595$, $q_{14-11} = 2.457$ and $w_{14-11} = 0.660$. In comparing these values with the critical values of Table 1. one finds that if the specified significance level is 0.10, all three test statistics reject a "no outliers" hypothesis. The statistics, v_{14-11} and q_{14-11} reject also at the 0.05 significance level, while w_{14-11} does not. This is to be expected since the probability plot demonstrates that the appropriate test statistic is v_{14-11} or q_{14-11} .

The slope of the line plotted in Figure 2 is about 1.1, giving an approximation of about 0.91/CQW = 0.91/1.076 = 0.845 for $\tilde{\sigma}_{14-11} = 0.842$. Note that again, the kth value has been weighted most heavily.

References

1. ANSCOMBE, F.J., "Rejection of Outliers," Technometrics, Vol. 2, No. 2 (1960), 123-147.

2. BARNETT, VIC, "The Study of Outliers: Purpose and Model," Appl. Statist. (1978), 242-50.

3. BARNETT, VIC and LEWIS, TOBY, Outliers in Statistical Data, John Wiley, New York, 1978.

4. CZUBER, E., Theorie der Beobachtungsfehler, Leipzig, 1981.

5. DANIEL, C., "Locating Outliers in Factorial Experiments," Technometrics, Vol. 2, No. 2 (1960), 149-156.

6. DIXON, W.J., "Analysis of Extreme Values," Ann. Math. Statist., 21, (1950), 488-506.

7. ENGELHARDT, M. and BAIN, L.J., "Some Complete and Censored Sampling Results for the Weibull or Extreme Value Distribution," Technometrics, Vol. 15, No. 3 (1973), 541-549.

8. _____ and _____, "Simplified Statistical Procedures for the Weibull or Extreme-Value Distribution," Technometrics, Vol. 19, No. 3 (1977), 323-331.

9. FERGUSON, T.S., "Rules for Rejection of Outliers," Rev. Int. Statist. Inst., Vol. 3, No. 1 (1961), 29-43.

10. GRUBBS, F.E., "Sample Criteria for Testing Outlying Observations," Ann. Math. Statist., Vol. 21, No. 1 (1950), 27-58.

11. _____, "Procedures for Detecting Outlying Observations," Technometrics, Vol. 11, No. 1 (1969), 1-21.

12. HARTER, H.L. and MOORE, A.H., "Maximum Likelihood Estimation, from Doubly Centored Samples, of the Parameters of the First Asymptotic Distribution of Extreme Values," J. Amer. Statist. Assoc., Vol. 63, No. 3 (1968), 889-901.

13. HARRIS, C.M. and KAYLAN, A., "Refinements on the Statistics of Recidivism Measurement," submitted for publication, 1979.

14. HAWKINS, D.M., "Comment on 'A New Statistic for Testing Suspected Outlines,'" Commun. in Statist. A6(5), 435-438.

15. KRUSKAL, W.H., FERGUSON, T.S., TUKEY, J.W., and GUMBEL, E.G., "Discussion of the Papers of Messrs. Anscombe and Daniel," Technometrics, Vol. 2, No. 2 (1960), 157-166.

16. LAWLESS, J.F., "A Prediction Problem Concerning Samples from the Exponential Distribution, with Application in Life Testing," Technometrics, Vol. 13, No. 4 (1971), 725-730.

17. _____, "Construction of Tolerance Bounds for the Extreme-Value and Weibull Distributions," Technometrics, Vol. 17, No. 2 (1975), 255-261.

18. MANN, N.R., "Results on Location and Scale Parameter Estimation with Application to the Extreme-Value Distribution," Aerospace Research Laboratories Report ARL 67-0023, Office of Aerospace Research, U.S. Air Force, Wright-Patterson Air Force Base, Ohio, 1967.

19. _____, "Tables for Obtaining the Best Linear Invariant Estimates of Parameters of the Weibull Distribution," Technometrics, Vol. 9, No. 4 (1967), 629-645.

20. _____, "Results on Statistical Estimation and Hypothesis Testing with Application to the Weibull and Extreme-Value Distributions, Aerospace Research Laboratories Report ARL 68-0068, Office of Aerospace Research, U.S. Air Force, Wright-Patterson Air Force Base, Ohio, 1968.

21. _____, "Use of Life-Test Data Analysis Methodology for Analyzing Undesirable Habitual Behavior," Serial T-406, The George Washington University Institute for Management Science and Engineering, Washington, D.C., 1979. Also in Proceedings of the 14th Annual Conference on Statistics, Computer Science and Operations Research, Cairo University, Cairo, Egypt, March, 1979.

22. _____, "Insights to Addictive Behavior Based on the Application of Life-Test Data-Analysis Methodology to Longitudinal Drug-Use Recidivism Data," to be submitted for publication.

23. _____ and FERTIG, K.W., "Tables for Obtaining Confidence Bounds and Tolerance Bounds Based on Best Linear Invariant Estimates of the Extreme-Value Distribution," Technometrics, Vol. 15, No. 1 (1973), 86-100.

24. _____ and _____, "Simplified Efficient Point and Interval Estimators for Weibull Parameters," Technometrics, Vol. 17, No. 3 (1975) 361-368.

25. _____ and _____, "A Goodness of Fit Test for the Two-Parameter Weibull Distribution Against Three-Parameter Weibull Alternatives," Technometrics, Vol. 17, No. 3 (1975), 237-245.

26. _____ and ROTHBERG, J., "Using Reliability Methodology to Think About Human Behavior," Presented at the Joint National TIMS/ORSA Meeting, Washington, D.C., 1980.

27. _____, SCHAFER, R.C. and SINGPURWALLA, N.D., Methods for Statistical Analysis of Reliability and Life Data, John Wiley, New York, 1974.

28. _____, SCHEUER, E.M. and FERTIG, K.W., "A New Goodness-of-Fit Test for Two-Parameter Weibull or Extreme-Value Distribution with Unknown Parameters," Commun. in Statist., Vol. 2, No. 5 (1973), 383-400.

29. PATNAIK, P.B., "The Non-Central χ^2 and F Distributions and Their Applications, Biometrika, Vol. 36, No. 2 (1949), 202-232.

30. SHAPIRO, S.S. and WILK, M.B., "An Analysis of Variance Test for Normality (Complete Samples)," Biometrika, Vol. 52, No. 4 (1965), 591-611.

31. TIKU, M.L., "A New Statistic for Testing for Suspected Outliers," Commun. in Statist., Vol. 4, No. 8 (1975), 737-752.

32. _____, "Rejoinder: Comment on a New Statistic for Testing Suspected Outliers," Commun. in Statist., A6(14) (1977), 1417-1422.

33. _____, "Robustness of MML Estimators Based on Censored Samples and Robust Test Statistics," J. Statistical Planning and Inference 4, (1980), 123-143.

34. _____, "Goodness of Fit Statistics Based on the Spacings of Complete or Censored Samples," Austral. J. Statist., Vol. 22 (1980), 260-275.

35. _____, and SINGH, M. "Testing Outliers in Multivariate Data. In Multivariate Distributions in Scientific Work. Eds.: G.P. Patil, C. Taillie and B. Baldessari. North Holland. To appear, 1980.

36. THOMAN, D.R., BAIN, L.J. and ANTLE, C.E., "Reliability and Tolerance Limits in the Weibull Distribution," Technometrics, Vol. 12, No. 2 (1970), 363-371.

37. TIETJEN, G.L. and MOORE, R.H., "Some Grubbs-type Statistics for the Detection of Outliers," Technometrics 14 (1972), 583-597.